Coláiste na Tríonóide, Baile Átha Cliath
Trinity College Dublin
Ollscoil Átha Cliath | The University of Dublin

Network Theory
Lecture 4.04

EEU45C09 / EEP55C09
Self Organising Technological Networks

Nicola Marchetti
nicola.marchetti@tcd.ie

BA model: Average Path Length

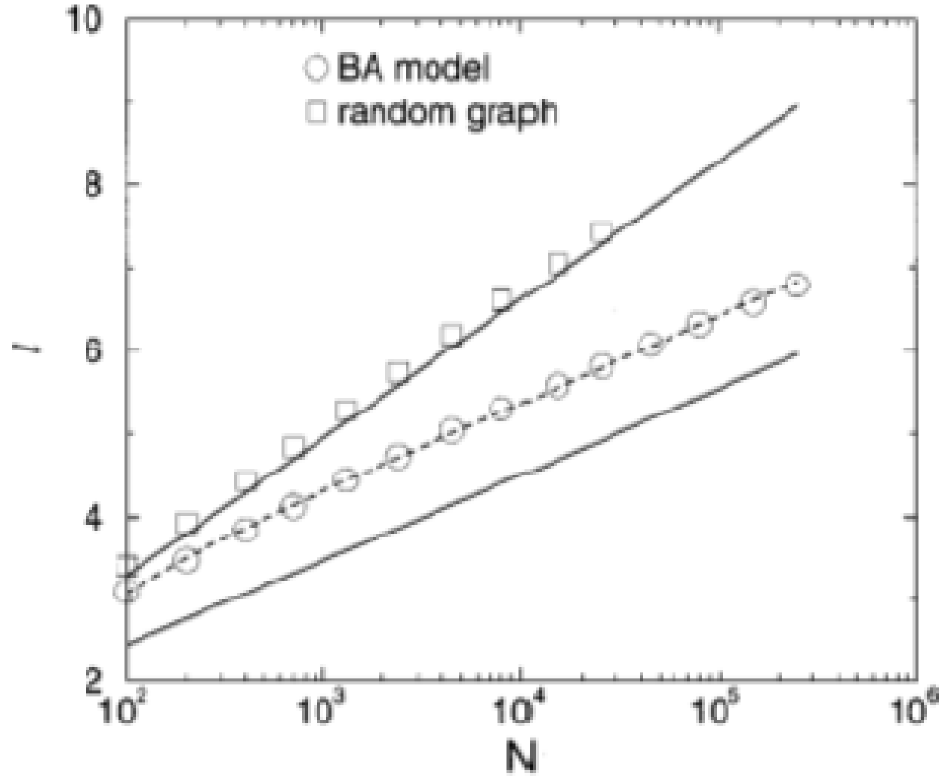

FIG. 23. Characteristic path length ℓ versus network size N in a Barabási-Albert (BA) network with $\langle k \rangle = 4$ (○), compared with a random graph of the same size and average degree generated with the algorithm described in Sec. III.A (□). The dashed line follows Eq. (94), and the solid lines represent Eq. (60) with $z_1 = \langle k \rangle$ and z_2 the numerically obtained number of next-nearest neighbors in the respective networks.

BA Model: Average Path Length

The average path length is lower for the BA model than an equivalent random graph for all values of N. It also grows slower than that for the random graph.

Numerically, Barabási and Albert have fit the average path length with a general logarithm form,

$$\langle l \rangle = A \log (N - B) + C.$$

Analytically, Bollobás and Riordan have indicated that the average path length in fact scales like,

$$\langle l \rangle \sim \frac{\ln N}{\ln \ln N}.$$

Scale-Free Network: Clustering coefficient

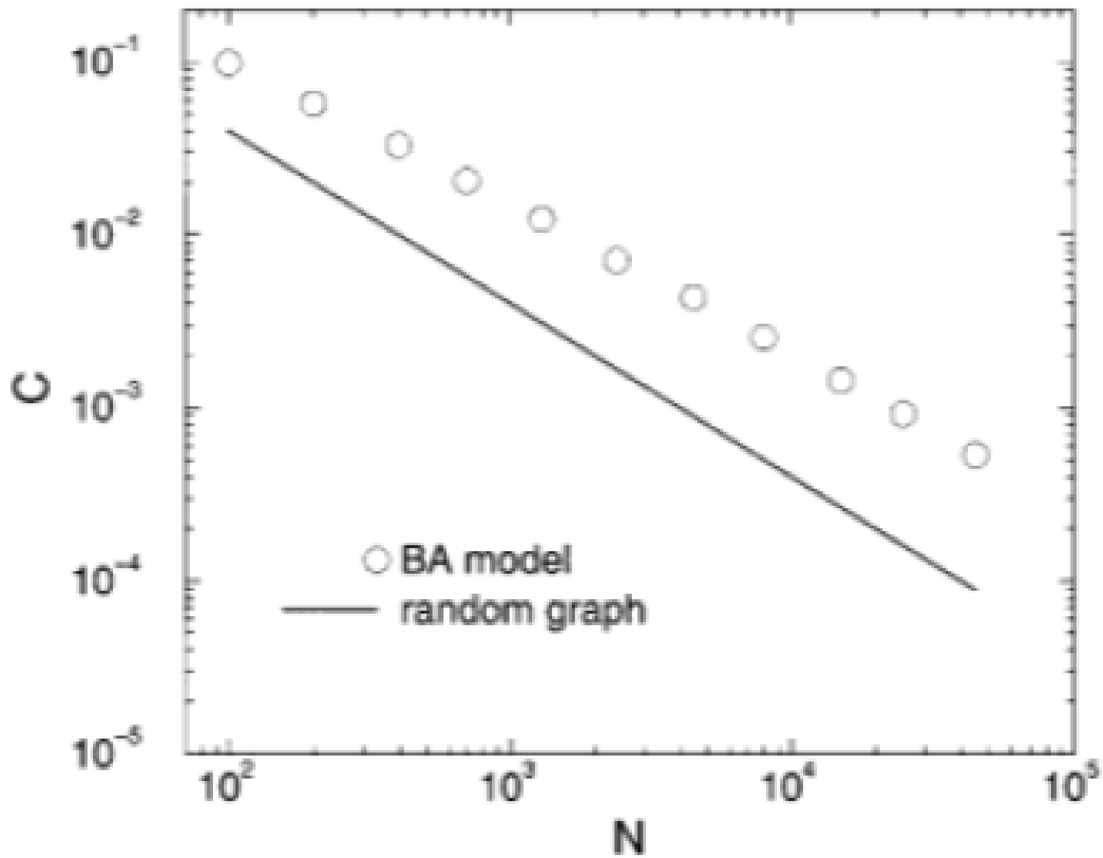

FIG. 24. Clustering coefficient versus size of the Barabási-Albert (BA) model with $\langle k \rangle = 4$, compared with the clustering coefficient of a random graph, $C_{rand} \simeq \langle k \rangle / N$.

Scale-Free Network: Clustering coefficient

In the previous slide, the clustering coefficient for the BA model is roughly five times that of the random graph.

The clustering coefficient of the BA model does decrease with system size. No analytical value for C, however numerically,

$$C \sim N^{-0.75}.$$

This is a slower decrease than the clustering coefficient of the random graph, $C = \langle k \rangle / N$

Figure 7: Caveat Emptor: Sometimes you see what you look for!

Searching for straight lines: A cautionary tale!

In the early 2000s, it became very popular to:

- gather data from a real-world network,
- calculate the degree distribution of the resulting graph,
- plot $P(k)$ using a log-scale on the axes,
- see a straight line and declare yourself discoverer of power-laws!

Searching for straight lines: A cautionary tale!

However, over a short interval of k values, other probability distributions can resemble power-laws.

One particular example is the log-normal distribution:

$$P(x) \sim \exp\left[-\frac{(\ln x - \mu)^2}{2\sigma^2}\right].$$

Searching for straight lines: A cautionary tale!

Over a limited interval of k, both functions resemble straight lines.

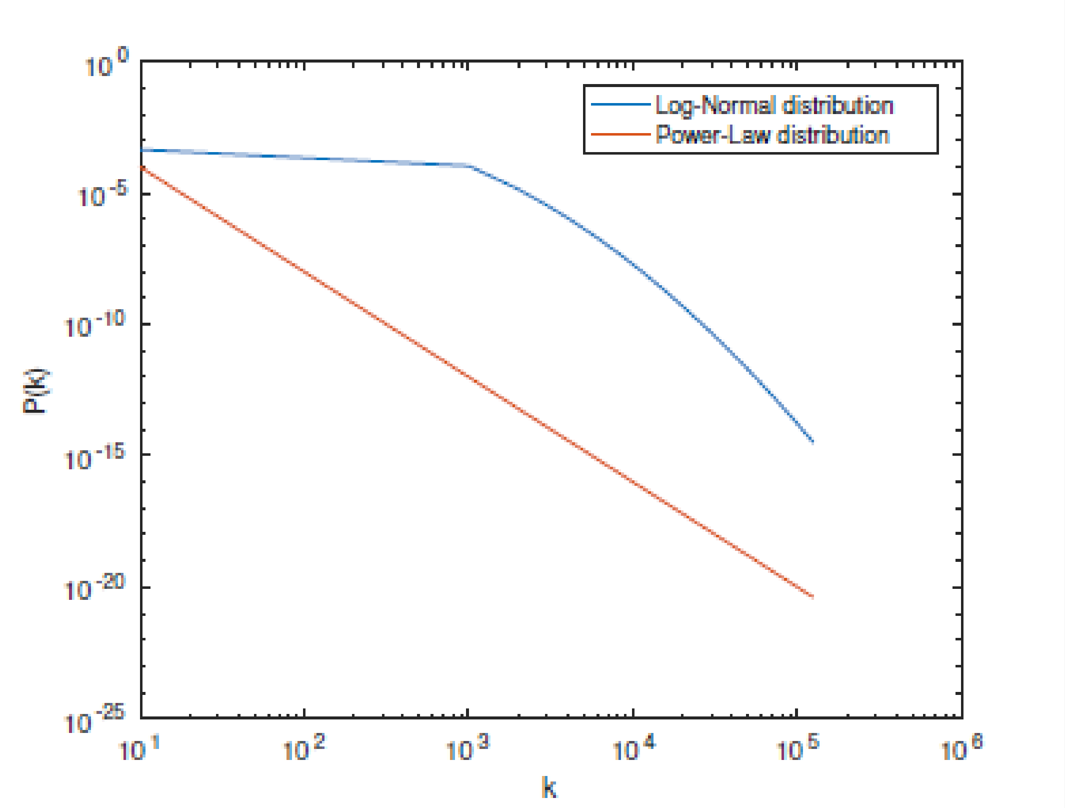

Figure 8: Log-log plot of a Power-Law distribution with $\gamma = 4$ and the Log-Normal distribution with mean $\mu = \ln(200)$ and standard deviation $\sigma = 1$.

Degree Centrality

Up to now, we have studied empirical and model networks. The degree and degree distribution have been useful to characterise these networks.

The degree is just one example of a network metric called Centrality.

Centrality asks the questions: Which are the most important or central nodes in the Network?

Degree Centrality

The simplest centrality measure in a network is the degree of each vertex, which is simply the number of edges connected to a given vertex.

Although simple, it can be very illuminating.

In social networks, individuals who have many connections to others might have more influence, more access to information or gossip and more prestige than those with a lower degree.

In citation networks of academic papers, the number of citations a paper receives from other papers (which is it's in-degree), is often used as a measure of whether a paper is influential or not.

Centrality

There are many possible definitions of importance and correspondingly many centrality measures in a network. e.g., we will look at the following centrality measures:

- Closeness centrality
- Betweenness centrality

Closeness centrality

Closeness centrality measures mean distance from a vertex to other vertices.

Suppose that d_{ij} is the length of the shortest path (geodesic path) through a network between two vertices i and j. The mean of this path length, to all vertices j in the network is

$$l_i = \frac{1}{N} \sum_j d_{ij}.$$

Therefore, l_i takes low values for vertices separated from others by only a small distance on average.

Closeness Centrality

A vertex with a small distance on average to other vertices has better access to information at those vertices or may have more influence on those vertices.

A person with low mean distance to others on social media, for example, will find their opinions reaching others quicker than someone with a larger mean distance.

Since l_i gives a low value for "more important" vertices in the network, it does not make for a good measure of centrality.

Closeness Centrality

Instead, the Closeness Centrality,

$$CC_i = \frac{1}{l_i} = \frac{N}{\sum_j d_{ij}},$$

can be used as a natural measure of centrality since it gives a large centrality to those vertices which are on average close to others.

Closeness centrality

Drawback: The values which the closeness centrality takes tend to span a very small range, from largest to smallest.

In graph that we have seen so far (ER, WS, BA model), the distance between vertices, d_{ij}, tends to be small on average with distance increasing logarithmically with system size.

Therefore, the ratio between the smallest distances (of order 1) and the largest (of order $\log N$)is small. These smallest and largest values provide upper and lower bounds on the average distance l_i, and therefore the range of values of l_i and CC_i is also small.

Closeness centrality

Typically, the values of CC_i spans a factor of five or less. Therefore, it becomes difficult to distinguish between central and less central nodes in practice.

The closeness centrality values tend to be cramped together and adjacent values may differ only by very small amounts.

This tendency can be seen in the graph of movie actors who have worked on films with each other. The actors can be ranked by their closeness centrality in this network. Using the data from IMDB, the highest closeness centrality of any actor is Christopher Lee with $CC = 0.4143$.

Closeness Centrality

In contrast, the lowest centrality actor is Leia Zanganeh, from Iran, with a closeness centrality of $CC = 0.1554$.

The ratio between the two is just 3.6 and about half a million other actors lie in between.

This leads to intermediate closeness centrality values being extremely close together and as a result, the rankings change frequently. This suggests that this is a dubious and unreliable measure of centrality.

Other measures of centrality, in particular the degree and eigenvector centrality, don't suffer from this problem because they have a wider dynamic range. Their centrality values, particularly of the leaders, also tend the be widely separated.

Closeness Centrality

A further complication occurs for networks with more than one component.

A connected component of an undirected graph is a subgraph in which every two vertices are connected by a path.

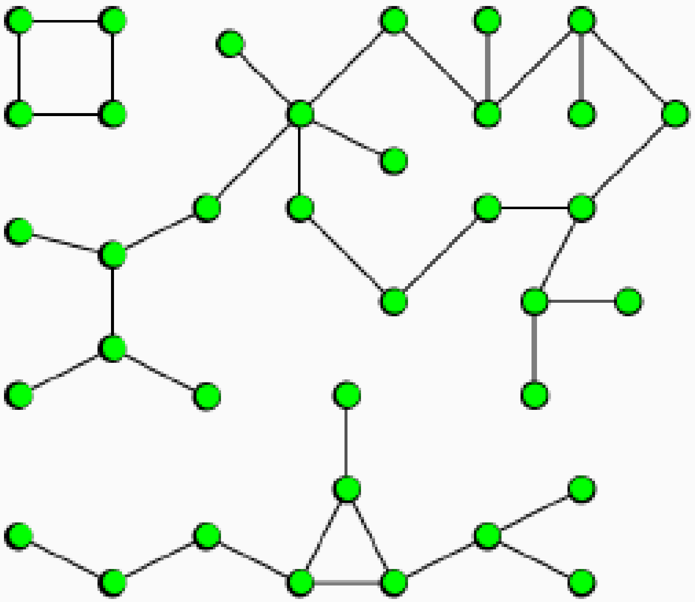

Figure 1: A graph with three different connected components.

Closeness Centrality

The path length between two vertices which are in two different components is defined to be infinite.

Therefore, l_i is infinite for all i in any network with more than one component. This leads to a closeness centrality $CC_i = 0$.

The most common workaround is to consider only those vertices in the same component. However, this gives smaller values of l_i to vertices in small components, giving them a larger closeness centrality than vertices in large components. This is undesirable since we consider vertices in small components to be less well connected than those in larger components.

A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, 1999.

M. Newman, *Networks: An Introduction.* **New York, NY, USA: Oxford University Press, Inc., 2010.**

R. Albert and A.-L. Barabási, "Statistical mechanics of complex networks," *Rev. Mod. Phys.*, vol. 74, pp. 47–97, Jan 2002.

Ackowledgement

- Dr Neal McBride, Arista Networks